UNTERSUCHUNGEN ZUR VALIDITÄT UND RELIABILITÄT DER DENVER-ENTWICKLUNGSSKALEN

Michael Roßberger

1979

Aus dem Institut für Soziale Pädiatrie
und Jugendmedizin der Universität München

Vorstand: Prof. Dr.med.Hellbrügge

UNTERSUCHUNGEN ZUR VALIDITÄT UND RELIABILITÄT DER DENVER-ENTWICKLUNGSSKALEN

Inaugural - Dissertation

zur Erlangung der Doktorwürde in der gesamten Medizin
an der Ludwig Maximimilians-Universität zu München

vorgelegt von

Michael Roßberger

aus

München

1979

UNIV.
LEIDEN
BIBL.

Mit Genehmigung der Medizinischen Fakultät der Universität München

Berichterstatter: Professor Dr.Th.Hellbrügge

Dekan: Professor Dr.W.Spann

Tag der mündlichen Prüfung: 31.Juli 1979

INHALTSVERZEICHNIS

	Seite
1. Einleitung	1
2. Literaturübersicht	3
2.1 Die Validität und Reliabilität eines Filtertests	3
2.2 Der Denver Developmental Screening Test	5
2.3 Die Denver-Entwicklungsskalen	8
2.3.1 Testaufbau und Testmaterial	8
2.3.2 Testdurchführung und Auswertung	8
2.3.3 Standardisierung	10
3. Problemstellung	11
4. Methodik	12
4.1 Beschreibung der Stichprobe	12
4.2 Beschreibung der Vergleichstests	14
4.2.1 Funktionelle Entwicklungsdiagnostik	14
4.2.2 Stanford-Binet Intelligenztest (S I T)	14
4.2.3 Neurologische Untersuchung	15
4.3 Prozedur und Berechnung	16
4.3.1 Gültigkeitsprüfung	16
4.3.2 Zuverlässigkeitsstudie	19
5. Ergebnisse	21
5.1 Gültigkeitsprüfung: Denver-Entwicklungsskalen und funktionelle Entwicklungsdiagnostik	21
5.2 Gültigkeitsprüfung: Denver-Entwicklungsskalen und Stanford-Binet Intelligenztest	22
5.3 Stabilitätsprüfung	26
6. Diskussion	28
6.1 Die Validität der Denver-Entwicklungsskalen	28
6.2 Die Reliabilität der Denver-Entwicklungsskalen	33

Seite

7. Zusammenfassung 34

8. Literaturverzeichnis 36

1. EINLEITUNG

Die gezielte Anwendung von sogenannten Screening- oder Filteruntersuchungen (JAHN 1966) ermöglicht es, eine Krankheit in ihrer präklinischen Phase zu entdecken und wirkungsvoller zu behandeln als in einem Stadium, in dem Symptome manifest und Schäden zum Teil irreversibel geworden sind (EPSTEIN 1964). Besonders in der Pädiatrie wurde auf die Notwendigkeit einer Präventivmedizin als Ergänzung der kurativen Medizin seit langem hingewiesen (HELLBRÜGGE 1965, 1971).

Die Definition "Screening", wie sie vom Regionalkomitee Europa der Weltgesundheitsorganisation anerkannt wurde, lautet: Versuch der frühzeitigen Erkennung unerkannter Krankheiten und Defekte durch die Anwendung von Tests, Untersuchungsmethoden oder rasch anwendbarer anderer Mittel. Screeningtests unterscheiden offenbar gut Personen, die vermutlich die Krankheit in sich tragen von solchen, die nicht von der Krankheit befallen sind. Ein Screeningtest ist seiner Intention nach keine Diagnose. Personen mit positivem oder verdächtigem Befund müssen ihren betreuenden Ärzten zur ausführlichen Diagnostik und zur notwendigen Behandlung überwiesen werden (eigene Übersetzung nach LESSLER 1974).

Die wichtigsten Kriterien für sinnvolle Screeninguntersuchungen lassen sich in folgende Punkte zusammenfassen (WILSON 1968):

1. Die Krankheit sollte ernst sein, das heißt die Belastung für den betreffenden Menschen und die Gesellschaft sollte offensichtlich sein, sowohl hinsichtlich des Leidens als auch der Kosten, wobei die Aufwendungen für die Screeninguntersuchung nicht höher sein sollten als für eine Behandlung, die erst in der symptomatischen Phase der Krankheit einsetzt.

2. Die Krankheit sollte ein aufdeckbares Frühstadium haben, für das diagnostische Möglichkeiten vorhanden sind, um echte Erkrankungen von Grenzfällen oder Normalbefunden zu unterscheiden.

3. Die Krankheit sollte behandelbar sein, und die Behandlung sollte beim Einsetzen im Frühstadium bessere Ergebnisse zeigen als im Spätstadium.

4. Der Sreeningtest zur Aufdeckung der Krankheit im Frühstadium sollte:

   a) ökonomisch sein,
   b) seine Gültigkeit oder Validität sowie
   c) seine Zuverlässigkeit oder Reliabilität sollten nachgewiesen sein,
   d) für den betreffenden Bevölkerungsteil sollte er akzeptabel sein.

In der vorliegenden Arbeit wurden die Denver-Entwicklungsskalen, ein Screeningtest für kindliche Entwicklungsstörungen im Altersbereich von zwei Wochen bis zu sechs Jahren, im Hinblick auf ihre Validität und Reliabilität untersucht.

2. LITERATURÜBERSICHT

2.1 Die Validität und Reliabilität eines Filtertests

Die Validität oder Gültigkeit eines Filtertests wurde von WILSON (1968) als die Häufigkeit beschrieben, mit der die Testergebnisse durch eine Diagnose bestätigt werden können. Es wird also die Genauigkeit geprüft, mit der ein Test kranke Personen von Gesunden trennt, wobei eine hundertprozentige Übereinstimmung selten erreicht wird. Letzteres erklärt sich schon aus der Natur des Filtertests, der eine einfachere Ausführung einer diagnostischen Untersuchung darstellt. Je weiter der Test von dieser entfernt ist, desto niedriger wird seine Gültigkeit auch sein. Die beste Methode der Gültigkeitsprüfung wäre eine Verlaufsbeobachtung, die die Wandlung der asymptomatischen zu symptomatischen Fällen registriert, was allerdings meist nur schwer durchführbar ist. KOLLER (1967) sprach in diesem Zusammenhang von direkter Validität im Gegensatz zur indirekten Validität, die sich ergibt, wenn ein Filtertest mit einer anderen diagnostischen Untersuchung verglichen wird, um die eventuell langdauernde Verlaufsbeobachtung zu ersetzen.

Objektiv läßt sich die Gültigkeit eines Filtertests mit den Begriffen Empfindlichkeit oder Sensitivität, Spezifität und diagnostischer Wert oder predictive value eines positiven bzw. negativen Testergebnisses beschreiben (VECCHIO 1966, WILSON 1968, PFLANZ 1973).

Die Empfindlichkeit gibt das Verhältnis der testpositiven Kranken zur Gesamtzahl der getesteten Kranken an, somit die Eignung des Tests Krankheit anzuzeigen. Eine Empfindlichkeit von 100 % bedeutet daher, daß bei allen Erkrankten ein positives Ergebnis gefunden wird.

Die Spezifität drückt das Verhältnis der testnegativen Gesunden zur Gesamtzahl der getesteten Gesunden aus.

Sie ist demnach ein Maß für die Tätigkeit eines Tests, die Symptomfreiheit eines Gesunden anzuzeigen.

BUCK und GART (1966) prägten die Begriffe copositivity und conegativity, wenn das Ergebnis eines Filtertests nicht mit einer durch die Verlaufsbeobachtung erhaltenen Diagnose verglichen wird, sondern nur mit einer dem Filtertest ähnlichen, jedoch ausführlicheren Untersuchung. Copositivity und conegativity sind also Begriffe, die die nach KOLLER (1967) als direkt bezeichnete Validität eines Filtertests beschreiben.

Der "predictive value" oder diagnostische Aussagewert eines positiven oder negativen Testergebnisses errechnet sich aus dem Verhältnis der richtig positiven bzw. negativen Ergebnisse des Filtertests zur Summe der richtig und falsch positiven Ergebnisse bzw. zur Summe der richtig und falsch negativen Ergebnisse der Diagnose.

Der "predictive value" gibt die Wahrscheinlichkeit an, mit der eine Erkrankung bei positivem Test vorliegt bzw. mit der der Untersuchte bei negativem Testausfall gesund ist (nach PFLANZ 1973).
KOLLER (1967) sprach in Zusammenhang mit der indirekten Validität von der positiven und negativen Korrektheit eines Filtertests, wobei die gleiche Berechnung wie beim "predictive value" verwendet wurde.

Bei einer Screeninguntersuchung sollte eine hohe Spezifität und damit eine hohe positive Korrektheit dafür sorgen, daß die Kosten für unnötige weitere Diagnostik nicht übermäßig zunehmen, andererseits sollen möglichst alle Symptomträger erfaßt werden. Das Probelm ist hierbei die inverse Beziehung zwischen Sensivität und Spezifität, die dazu zwingt, für die Entscheidungsgrenze zwischen Krankheit und Gesundheit im Hinblick auf die jeweilige Fragestellung einen Kompromiß zu finden (WERNER 1977). Als Möglichkeit der Verbesserung des "predictive value"

eines Tests werden die Steigerung der Prävalenz der Krankheit durch Präselektion sowie Erhöhung der Spezifität des Tests, wobei eventuell eine Minderung der Empfindlichkeit in Kauf genommen werden muß, genannt (OELLERICH u. HAECKEL 1978).

Als Zuverlässigkeit oder Reliabilität eines Filtertests wird die Reproduzierbarkeit seiner Ergebnisse bezeichnet. Sie bezieht sich sowohl auf die Übereinstimmung der Ergebnisse, die von zwei unabhängigen Untersuchern erhoben werden (Objektivität eines Tests), als auch auf die Stabilität eines Tests (nach PFLANZ 1973). Der diagnostische Wert eines Tests wird durch seine Zuverlässigkeit entscheidend mitbeeinflußt (EGGSTEIN 1975).

2.2 Der Denver Developmental Screening Test

Zur weiteren Einführung in die Problemstellung soll nun der zu untersuchende Test, die Denver-Entwicklungsskalen, im Hinblick auf seine Standardisierung, seine bisher beschriebene Validität und Reliabilität sowie seine praktische Durchführung vorgestellt werden.

Die Denver-Entwicklungsskalen stellen die deutsche Standardisierung und Bearbeitung des Denver Developmental Screening Test dar, der 1967 von FRANKENBURG und DODDS veröffentlicht wurde.

Der Denver Developmental Screening Test ist als Screeningtest zur Auffindung kindlicher Entwicklungsstörungen im Altersbereich von zwei Wochen bis zu sechs Jahren konstruiert. Die Intention dieses Testes liegt nach Aussagen der Autoren nicht darin, IQ- oder EQ-Werte festzustellen und Diagnosen zu liefern, sondern Hinweise zu geben für Auffälligkeiten des kindlichen Verhaltens, die genauerer Untersuchung und Abklärung bedürfen.

Aus Entwicklungs- und Vorschulintelligenztests wurden insgesamt 240 Merkmale (Items) ausgewählt, die einfach anzuwenden und auszuwerten waren.

Folgende Tests lieferten die Items:

1) Bayley Infant Scales of Development
2) Cattell Infant Scales of Development
3) Gesell Developmental Schedules
4) Griffith Developmental Scales
5) Hetzer-Wolf Baby Scales from the Vienna Tests
6) Merril-Palmer Scale of Mental Tests
7) Revised Yale Developmental Schedules
8) Stanford-Binet Intelligence Scale
9) Lincoln Oseretzky Motor Development Scale
10) Communication Evaluation Chart
11) Vineland Social Maturity Scale
12) The California Developmental Charts

Zunächst wurden 200 Säuglinge und Vorschulkinder versuchsweise mit diesen Merkmalen getestet. 105 Merkmale wurden für die eigentliche Standardisierung ausgewählt, da sie sich als handlich in der Durchführung erwiesen, und sie nur eine geringe Altersstreuung zeigten. Die umfassende Standardisierung der Merkmale war nach Ansicht der Autoren nötig, weil die Altersangaben für die einzelnen Items in verschiedenen Tests stark variierten. Die Stichprobe der Standardisierung in Denver bestand aus 543 Jungen und 493 Mädchen (N = 1036) im Alter von zwei Wochen bis 6,4 Jahren. Ausgeschlossen wurden Kinder mit ausgeprägten physischen Defekten wie Hör- und Sehschäden sowie Risikokinder wie Frühgeburten und Zangengeburten. Adoptierte Kinder und Zwillinge wurden ebenfalls nicht in die Stichprobe aufgenommen. Die Auswahl der Kinder entsprach weitgehend der ethnischen und soziologischen Zusammensetzung der Bevölkerung in Denver.

Es wurde eine Einteilung in 72 Altersgruppen vorgenommen und für jede Altersgruppe der Zeitpunkt berechnet, zu dem 25%, 50%, 75% und 90% der Kinder die jeweiligen Merkmale erfüllten. Bei einer vergleichenden Betrachtung der Ergebnisse ergab sich, daß in den ersten beiden Lebensjahren kein bemerkenswerter Unterschied bei den Kindern aus verschiedenen sozio-ökonomischen Schichten bestand. Nach dem zweiten Lebensjahr konnten Kinder, deren Väter gehobenere Berufe ausübten, einige Merkmale im Sprachbereich eher ausführen. Zwischen den Geschlechtern wurden einige geringe, nicht signifikante Unterschiede hinsichtlich der Erfüllung der Aufgaben festgestellt.

Bei der Prüfung der Validität des Denver Developmental Screening Tests an 263 Kindern mit dem Stanford-Binet-Test, Revised Yale Developmental Schedule, Cattell und Revised Bayley Infant Scale als Vergleichstest ergab sich eine copositivity von 100% und eine connegativity von 77% beim Vergleich mit dem Stanford-Binet-Test. Bei den anderen Tests betrugen die Werte entsprechend der genannten Reihenfolge 97% bzw. 58%, 57% bzw. 94% und 29% bzw. 88% (FRANKENBURG et al. 1971).

FRANKENBURG et al. entschlossen sich 1971 die Bewertungskriterien des Tests zu revidieren, da sich in der Praxis gezeigt hatte, daß der Anteil der falsch positiven Ergebnisse zu hoch war. Der Anteil der falsch positiven Resultate bei einem Vergleich des Denver Tests mit dem Stanford-Binet-Test oder der Revised Bayley Infant Scale mit Hilfe der Testergebnisse von 237 Kindern im Alter von ein Monat bis zu sechs Jahren konnte von 75% der Gesamtzahl der Testpositiven auf 57% gesenkt werden. Als Erklärung für den noch immer hohen Anteil trotz Revision der Bewertungskriterien nahmen die Autoren an, daß Merkmale im Denver Developmental Screening Test enthalten seien, die der Stanford-Binet Test nicht messe. Bei der Prüfung der Stabilität ergab sich eine allgemeine Übereinstimmung von 97% der aus zweimaliger Testung im Abstand von einer Woche erhaltenen Ergebnisse von 186 Kindern (FRANKENBURG et al. 1971).

2.3 Die Denver-Entwicklungsskalen

2.3.1 Testaufbau und Testmaterial

Die 105 Items der Denver Entwicklungsskalen sind auf einem Testblatt zusammengefaßt und entsprechen den Daten der Eichstichprobe entlang einer Alterslinie. Diese Alterslinie ist bis zum zweiten Lebensjahr in Monatsabschnitte unterteilt, bis zum fünften Lebensjahr in 3-Monats-Abschnitte. Die Merkmale wurden inhaltlich in vier Bereiche eingeordnet (siehe Testblatt im Anhang): sozialer Kontakt, Feinmotorik und Adaptation, Sprache, Grobmotorik. Die Bedeutung dieser vier Bereiche hinsichtlich der Beurteilung eines Kindes sollte nicht überbewertet werden, da die Zuordnung nach Angaben der Autoren weitgehend "a priori" erfolgte. Jedes Merkmal wird auf dem Testblatt durch einen waagrechten Balken graphisch dargestellt, auf dem markiert ist, wann 25%, 50%, 75% und 90% der Kinder der Eichstichprobe dieses Merkmal erfüllten.

Aufgaben, die von der Erziehungsperson als durchgeführt berichtet werden, sind mit einem "+" gekennzeichnet. Fußnoten in den Testbalken verweisen auf die Erklärungen auf der Rückseite des Testbogens.

Als Testmaterial wird verwendet: rote Wolle, Rosinen, Klapper, 10 farbige Klötzchen, eine kleine Glocke, ein Tennisball, Bleistift, Papier, eine kleine Glasflasche mit enger Öffnung. Die Rückseite des Testblattes enthält folgende Abbildungen: Katze, Vogel, Pferd, Hund, Mensch, Kreis, zwei unterschiedlich lange senkrechte Linien, Kreuz, Quadrat.

2.3.2 Testdurchführung und Auswertung

Zur Durchführung des Tests wird das Alter des betreffenden Kindes auf der Alterslinie aufgesucht und durch einen zu dieser Linie geführten senkrechten Strich über das ganze Test-

blatt festgehalten. Bei Frühgeborenen wird das chronologische Alter korrigiert, indem man das Alter um das Ausmaß der Frühgeburtlichkeit reduziert. Alle Merkmale, die von dem senkrechten Strich geschnitten werden, sollen geprüft werden. Doch sollen auch Aufgaben links und rechts der Linie gegeben werden, bis bei einer Entwicklungsverzögerung eine durchgehende Reihe von bestandenen Aufgaben erhalten wird, oder bei einem Entwicklungsvorsprung eine Reihe von Aufgaben, die nicht mehr als bestanden bezeichnet werden können. Die Durchführung und Bewertung der Merkmale werden im Testmaterial detailliert beschrieben.

Die Items werden als bestanden oder als Fehler gekennzeichnet. Wenn das Kind- trotz wiederholter Versuche - nicht bereit ist, eine bestimmte Aufgabe durchzuführen, wird diese als Verweigerung markiert.

Zur Auswertung des Tests werden sämtliche Fehlermerkmale, die von 90 % der Gleichaltrigen erfüllt werden, also alle Fehlermerkmale, die links der senkrechten Altersmarkierung liegen, für jeden Bereich summiert. Die Gesamtbewertung erfolgt nach den Kategorien: normal, fraglich und abnorm.

Als "abnorm" wird ein Testergebnis bezeichnet, wenn in mindestens zwei Bereichen zwei oder mehr Fehler gezählt werden, oder wenn in einem Bereich zwei Fehler gefunden werden und ein anderer Bereich einen Fehler und kein "bestanden" in dem Feld aufweist, das von der senkrechten Alterslinie geschnitten wird.

Als "fraglich" wird ein Ergebnis bezeichnet, wenn in einem Bereich mindestens zwei Fehler gezählt werden oder in einem oder mehreren Bereichen ein Fehler und kein "bestanden", das von der Alterslinie geschnitten wird.

"Normal" ist ein Ergebnis, wenn weder die Bedingungen für "fraglich" noch für "abnorm" erfüllt werden.

Als zu diesem Zeitpunkt untestbar wird ein Kind bezeichnet, das so viele Verweigerungen gezeigt hat, daß, wenn diese Fehler wären, der Test als "abnorm" oder "fraglich" gelten würde. In diesen Fällen wird ein erneutes Screening nach einigen Wochen empfohlen.

2.3.3 Standardisierung

Da verschiedene Kultureinflüsse eine unterschiedliche Entwicklung im Kindesalter bedingen, war es sinnvoll, eine Standardisierung der Merkmale des Denver Developmental Screening Tests entsprechend den Bedingungen in der Bundesrepublik Deutschland vorzunehmen. Eine derartige Standardisierung wurde von FLEHMIG durchgeführt und 1973 wurde als Ergebnis ein deutschsprachiges Testblatt gemäß den erhaltenen Daten vorgelegt.

Eine eingehende Beschreibung der Standardisierung liegt bis jetzt nicht vor. Aus einem Übersichtsreferat, das FLEHMIG 1975 veröffentlichte, ergeben sich folgende grobe Daten: Es wurden 1455 Kinder im Alter zwischen zwei Wochen und sechs Jahren in Hamburg und Düsseldorf getestet. Der Anteil der Mädchen lag bei 46.5 %. Die sozio-ökonomischen Daten wurden berücksichtigt. Die gesamte Stichprobe wurde in 36 Altersgruppen mit je 40 Kinder aufgeteilt und entsprechend der amerikanischen Standardisierung für jede Aufgabe das Alter bestimmt, in dem 25 %, 50 %, 75 % und 90 % der Kinder diese Aufgabe erfüllt hatten. Es wurden 1240 Kinder in Kliniken und 215 Kinder in Kindergärten getestet. 35,9 % der Mütter waren berufstätig, das Durchschnittsalter der Mütter lag bei 27,9 Jahren, das der Väter bei 30,6 Jahren. Es fand sich ein Mittelwert von 1,7 Kindern pro Haushalt.

3. PROBLEMSTELLUNG

In der vorliegenden Arbeit sollte vor allem die indirekte Validität, wie sie oben definiert wurde, der Denver Entwicklungsskalen bei der Entdeckung von Kindern im Alter von drei bis sechs Jahren mit einem Entwicklungsrückstand im Intelligenzbereich untersucht werden. Als Vergleichstest diente der Stanford-Binet Test in der Bearbeitung von LÜCKERT.

Entsprechend den beschriebenen Erfahrungen bei der Validitätsstudie des Denver Developmental Screening Tests (FRANKENBURG 1971) sollte untersucht werden, ob der hohe Prozentsatz an falsch-positiven Ergebnissen durch den Anteil neurologischer Merkmale erklärt werden kann, die zwar in den Denver-Entwicklungsskalen enthalten sind, jedoch nicht in einem herkömmlichen Intelligenztest wie dem Stanford-Binet Test. Eine vergleichbare Untergruppe wurde zur Klärung dieses Problems zusätzlich neurologisch untersucht.

Zu einer vorläufigen Studie, die erste Hinweise für eine ausführlichere Validitätsuntersuchung geben sollte, sollte die Testung einer Gruppe von Kindern bis zu drei Jahren dienen, die mit den Denver-Entwicklungsskalen und der Münchener funktionellen Entwicklungsdiagnostik (HELLBRÜGGE et al. 1971, 1974) untersucht wurden.

Außerdem sollte geprüft werden, ob sich die Bewertungskriterien der Denver Entwicklungsskalen, wie sie im Testmanual angegeben sind, modifizieren lassen zur Steigerung des diagnostischen Wertes eines positiven Testergebnisses. Die Optimierung dieses Wertes ist vor allem bei einem Screeningtest vorrangig, wie oben ausgeführt wurde.

Zur Prüfung der Stabilität als Teil der Reliabilität eines Screeningtests sollte eine Gruppe von Kindern aus der Validitätsstudie zwischen drei und sechs Jahren in einem bestimmten zeitlichen Abstand getestet und die Ergebnisse verglichen werden.

4. METHODIK

4.1 Beschreibung der Stichprobe

Insgesamt wurden 173 Kinder aus der Ambulanz des Kinderzentrums München (34 %), der Ambulanz der Universitätskinderklinik München (20 %), aus heilpädagogischen Kindergärten (27 %) und aus einem Privatkindergarten (19 %) getestet.

25 Kinder waren jünger als drei Jahre und wurden in einer eigenen Untersuchungsgruppe zusammengefaßt; die Verteilung nach Alter und Geschlecht ist in Tabelle 1 zusammengestellt.

Eine zweite Gruppe bestand aus 148 Kindern zwischen drei und sechs Jahren mit einer Alters- und Geschlechtsverteilung entsprechend Tabelle 2. Hieraus wurde eine Untergruppe von 111 Kindern gebildet, die auch neurologisch untersucht wurden und die in Tabelle 3 beschrieben wird.

Die Stichprobe zur Stabilitätsprüfung der Denver Entwicklungsskalen bestand aus einer Untergruppe von 38 Kindern aus der Gesamtgruppe mit 148 Kindern mit einer Alters- und Gechlechtsverteilung, wie sie in Tabelle 4 dargestellt wird.

	Alter in Jahren			
	1	1	2	
männlich	7	5	8	20
weiblich	4	1	0	5
	11	6	8	25

Tab. 1: Alters- und Geschlechtsverteilung der Kinder unter drei Jahren (N = 25)

	Alter in Jahren 3	4	5	
männlich	17	32	37	86
weiblich	15	14	33	62
	32	46	70	148

Tab. 2: Alters- und Geschlechtsverteilung der Kinder über drei Jahren (N = 148)

	Alter in Jahren 3	4	5	
männlich	12	24	30	66
weiblich	12	10	23	45
	24	34	53	111

Tab. 3: Alters- und Geschlechtsverteilung der Kinder über drei Jahren, die zusätzlich neurologisch untersucht wurden (N = 111)

	Alter in Jahren 3	4	5	
männlich	2	5	5	12
weiblich	2	9	15	26
	4	14	20	38

Tab. 4: Alters- und Geschlechtsverteilung der Kinder über drei Jahren, die zur Stabilitätsprüfung zweimal untersucht wurden (N = 38)

4.2 Beschreibung der Vergleichstests

4.2.1 Funktionelle Entwicklungsdiagnostik

Die funktionelle Entwicklungsdiagnostik für das erste und zweite Lebensjahr wurde von HELLBRÜGGE et al. (1971, 1974) als diagnostisches Instrument zur differenzierten Erfassung frühkindlicher Entwicklung im Hinblick auf die wesentlichen Funktionsbereiche entwickelt. Sie soll zugleich Grundlage für eine funktionelle Entwicklungstherapie sein (SCHAMBERGER, 1975). Für das erste und zweite Lebensjahr wurden acht Funktionsbereiche vorgesehen, nämlich Krabbeln, Sitzen, Laufen, Greifen, Perzeption, Lautäußerung und Sprechen, Sprachverständnis und Sozialentwicklung. Bei der Standardisierung wurde als Mindestnorm für jede Aufgabe derjenige Lebensmonat festgesetzt, in dem 90 % der Kinder der entsprechenden Altersstufe diese erfüllen konnten. Für jeden Funktionsbereich ergibt die Untersuchung eines Kindes ein Entwicklungsalter. Die graphische Darstellung der einzelnen Entwicklungsalter ergibt das Entwicklungsprofil (SCHAMBERGER 1975). Daten über Gültigkeit und Zuverlässigkeit lagen zum Zeitpunkt dieser Untersuchung noch nicht vor.

4.2.2 Stanford-Binet Intelligenztest (S I T)

Der verwendete Stanford-Binet Intelligenztest stellt eine 1957 von LÜCKERT veröffentlichte, für deutsche Verhältnisse standardisierte Form des 1937 und 1960 von TERMAN veröffentlichten Intelligenztestes für das 3. - 14. Lebensjahr dar. Die Aufgaben waren von den Autoren nach Testung von 842 Kindern im Alter von drei bis fünf Jahren, von 2638 Kindern älterer Jahrgänge geeicht worden. Regionale Unterschiede und Schichtung wurden berücksichtigt. Zur Standardisierung mußte eine Aufgabe von 78 % der Kinder einer Altersstufe erfüllt worden sein. Für die Drei- und Vierjährigen entspricht eine Aufgabe einem Monat, für die Älteren zwei Monate. Aus dem Quotient

Intelligenzalter und Lebensalter multipliziert mit 100 wird der Intelligenzquotient (IQ) berechnet. Angaben über Zuverlässigkeit und Gültigkeit liegen nicht vor.

In Ermangelung eines anderen Tests für die Altersstufe von drei bis sechs Jahren und wegen seiner weiten Verbreitung, die wohl auf seine praktische Brauchbarkeit zurückzuführen ist, wurde dieser Test zur Gültigkeitsprüfung der Denver Entwicklungsskalen in dieser Untersuchung verwendet.

4.2.3 Neurologische Untersuchung

Die bei einem Teil der Kinder zusätzlich angewandte neurologische Untersuchung, die von Frau Dr.B.Ohrt und Frau cand.med.G.Trinklein von der Universitätskinderklinik München zusammengestellt und durchgeführt wurde, entsprach weitgehend einem 1970 von TOUWEN und PRECHTL veröffentlichten Untersuchungsschema und enthielt folgende Merkmale:

Auffälligkeit der Kopfform, Symmetrie der Sitzhaltung, Widerstand gegen passive Bewegungen und Extensibilität in den Hals-, Schulter-, Wirbel-, Ellenbogen-, Hand-, Hüft-, Knie-, Finger- und Fußgelenken, Registrierung der Häufigkeit, des Tempos, der Ausführung der Zielgerichtetheit von spontanen Bewegungen im Sitzen und Stehen, Konsistenz, Kraft und Masse der Muskulatur, Prüfung propriorezeptiver Reflexe, des Babinskireflexes, der Oberflächen- und Tiefensensibilität, des 5., 7. und 11. Hirnnerven, des Gehörs mit Flüstername, Papierknistern und Glocke; Lautbestand der Sprache, Sprachfluß, Satzlänge, Sprachverständnis; Sehprüfung mit Zeigen einer Karte, eines Bildchens, Ausschluß eines latenten oder manifesten Strabismus, eines Nystagmus, grobe Beurteilung des Gesichtsfeldes, Pupillenreaktion auf Licht und Konvergenz; Haltung im Stehen, Fußabstand, Haltung mit ausgestreckten Armen, Auftreten unwillkürlicher Bewegungen im Stehen; Prüfung der Koordination bei Diadochokinese, beim Finger-Nase-Versuch, Finger-Finger-Versuch, Finger-Oppo-

sition, Öffnen und Schließen der Hände, Klopfen, Ausschluß assoziierter Bewegungen, Beurteilung der Hampelmannbewegung, der Balance mit offenen und geschlossenen Augen; Beschreibung der Haltung beim Gehen, der Ausführung des Fußabstands, des Gehens auf einer Linie, des Zehengangs, des Fersenganges, des Stehens auf einem Bein, der dabei auftretenden Balancebewegungen, Hüpfen auf einem Bein, Rennen; Beschreibung der Haltung im Liegen, die Durchführung des Fuß-Fingertippens, des Knie-Hackenversuches, des Aufsetzens, des Seit- und Langsitzes.

Das Ergebnis der Untersuchung wurde als unauffällig oder auffällig und damit therapiebedürftig beschrieben und eine klinische Diagnose erstellt. Als auffällig wurde das Gesamtergebnis angesehen, wenn ein Kind bei mindestens drei von folgenden sieben Kriterien auffällig war:

1. Deutliche Retardation der motorischen Entwicklung
2. Haltungsasymmetrie oder pathologische Haltung
3. Deutliche Veränderung des Muskeltonus und der Muskulatur
4. Koordinationsstörungen der Grob- und/oder Feinmotorik
5. Gleichgewichtsstörungen
6. Persistieren tonischer Reaktionen oder pathologischer, unwillkürlicher Bewegungen
7. Funktionsstörungen der Sinnesorgane und Sensibilitätsstörungen.

4.3 Prozedur und Berechnung

4.3.1 Gültigkeitsprüfung

Mit den Denver Entwicklungsskalen wurden alle 173 Kinder dieser Untersuchungen getestet. In keinem längeren Abstand als einer Woche wurden davon 25 Kinder zusätzlich mit der funktionellen Entwicklungsdiagnostik untersucht, nämlich alle Kinder unter

drei Jahren; 148 Kinder im Alter zwischen drei und sechs Jahren wurden mit dem Stanford-Binet Intelligenztest untersucht. Diese beiden Testuntersuchungen wurden von Psychologen und Ärzten des Kinderzentrums durchgeführt. Von der Gruppe der 148 Kinder wurden 111 zusätzlich neurologisch untersucht. Die Ergebnisse des jeweils vorhergehenden Tests waren den Untersuchern nicht bekannt. Entsprechend der Testanweisung der Denver Entwicklungsskalen wurden die getesteten Kinder in die Gruppen mit dem Testergebnis "abnorm", "fraglich", "normal" eingeteilt. Aus den Resultaten der funktionellen Entwicklungsdiagnostik wurde ein durchschnittliches Entwicklungsalter und daraus der Entwicklungsquotient (EQ) errechnet, um ein Gesamtergebnis für den Vergleich mit dem Ergebnis der Denver Entwicklungsskalen zu erhalten. Die Einteilung der EQ- bzw. IQ-Werte des Stanford-Binet Tests in vier Gruppierungen erfolgte modifiziert nach der Einteilung von WECHSLER (1956).

Das Ergebnis der neurologischen Untersuchung wurde als "auffällig" oder "unauffällig" angegeben entsprechend den oben genannten Kriterien.

Die Beziehungen zwischen Filtertestresultat und Diagnose kann am deutlichsten an Hand einer Vierfeldertafel (Abb. 1) gezeigt werden. Kinder, die übereinstimmend als krank, durch Filtertest und Vergleichstest bzw. Diagnose klassifiziert werden, werden in Feld A zusammengefaßt, übereinstimmend als gesund bezeichnete in Feld D. Falsch positive und falsch negative Befunde des Tests erscheinen in den Feldern B bzw. C. Die allgemeine prozentuale Übereinstimmung, als Summe von A und D dividiert durch A + B + C + D, stellt ein wenig differenziertes Maß dar, weil sie von den Randverteilungen stark abhängig ist. Informativer ist die Berechnung der Empfindlichkeit oder Sensitivität und der Spezifität eines Tests.

		Diagnose		
		Kranke	Gesunde	Summe
Filtertest	positiv	A	B	A+B
	negativ	C	D	C+D
		A+C	B+D	A+B+C+D

Abb. 1: Vierfeldertafel (nach PFLANZ 1973)

A = richtig Positive
B = falsch Positive
C = falsch Negative
D = richtig Negative

Die Sensitivität bezeichnet die Eignung eines Filtertests, Personen mit der fraglichen Krankheit zu erfassen. Aus der Vierfeldertafel läßt sich errechnen: Sensitivität in Prozent = 100 A/A+C. Die Eignung des Tests, möglichst alle gesunden Personen zu erfassen, wird durch die Spezifität beschrieben: Spezifität in Prozent = 100 D/B+D (PFLANZ 1973).

BUCK und GART (1966) prägten die Begriffe "copositivity" und "conegativity", wenn die Diagnose nicht durch Verlaufsbeobachtung, sondern nur durch eine dem Filtertest ähnliche lediglich ausführlicherer Untersuchung gestellt wird. In dieser Arbeit soll diese Überlegung berücksichtigt werden und die Sensitivität als positive Übereinstimmung und die Spezifität als negative Übereinstimmung bezeichnet werden.

Neben der Rate der falsch positiven Befunde wurde von KOLLER (1967) eine "positive Korrektheit" eines Filtertests angege-

ben, die sich in Prozent berechnet: 100 A/A+B. Die negative Korrektheit ist aus der Formel 100 D/C+D zu berechnen. Die positive oder negative Korrektheit wird auch als diagnostischer Aussagewert oder predictive value eines positiven oder negativen Befundes bezeichnet, wie oben ausgeführt wurde.

Zur Berechnung der eben angegebenen Prozentwerte wurden Vierfeldertafeln aufgestellt, wobei die Ergebnisse "fraglich" der Denver Entwicklungsskalen zunächst zu den "Abnormen" und dann zu den "Normalen" gezählt wurden, so daß sich jeweils zwei Gruppen ergaben. Die Ergebnisse der funktionellen Entwicklungsdiagnostik und des Stanford-Binet Tests wurden so zusammengefaßt, daß aus der Vierfeldertafel jeweils die positive und negative Übereinstimmung bzw. Korrektheit der Denver Entwicklungsskalen bei der Erfassung von Kindern mit einem EQ bzw. IQ unter 70, unter 80 und unter 90 berechnet werden konnte.

Die Ergebnisse der neurologischen Untersuchung wurden bei der entsprechenden Gruppe von Kindern insofern berücksichtigt, daß wenn sich bei einem Kind ein auffälliger neurologischer Befund ergab, das positive Testergebnis der Denver Entwicklungsskalen auch dann als richtig positiv galt, wenn keine Übereinstimmung mit dem Ergebnis des Intelligenztests bestand. Der umgekehrte Fall wurde in diesem Zusammenhang nicht bewertet.

4.3.2 Zuverlässigkeitsstudie

38 der 173 Kinder wurden zweimal im Abstand von einer Woche mit dem Denver Entwicklungstest untersucht. Erfahrungsgemäß, wenn auch empirisch nicht ausreichend begründet, genügt dieser Zeitraum, um Lern- und Erinnerungseffekte auszuschalten (PFLANZ 1973); bei einem längeren Abstand als vier Wochen würde die Veränderung des erfragten Merkmals durch die fortschreitende Entwicklung des Kindes an Bedeutung gewinnen. Die Auswertung des ersten und zweiten Tests erfolgte erst, nachdem

der zweite abgeschlossen war, so daß der Testleiter keine Kenntnis von der Gesamtbewertung des ersten Tests hatte.

Die Beurteilung der Zuverlässigkeit eines Tests erfolgte ebenfalls mit Hilfe einer Vierfeldertafel. Ein Maß für die Übereinstimmung ergibt sich als Prozentwert aus der Formel: Stabilität = (A+D) x 100/A+B+C+D (nach PFLANZ 1973). Um diesen Wert nicht trügerisch hoch werden zu lassen, wenn er vor allem durch die Zahl in Feld D gestützt wird, schlugen BLOHMKE und MAIER (zitiert nach PFLANZ 1973) als "Wiederholbarkeit" eines Tests folgende Berechnung vor: A x 100/A+B+C.

5. ERGEBNISSE

5.1 Gültigkeitsprüfung: Denver Entwicklungsskalen und funktionelle Entwicklungsdiagnostik

In Tabelle 5 sind die Testergebnisse der Gruppe (N = 25) von Kindern im Alter zwischen drei Monaten und drei Jahren, die mit den Denver Entwicklungsskalen und der funktionellen Entwicklungsdiagnostik untersucht wurden, zusammengefaßt. Die Angaben für die Berechnung der positiven und negativen Übereinstimmung bzw. Korrektheit ergaben sich aus Vierfeldertafeln, wobei die EQ-Stufen und die Bewertungen der Denver Entwicklungsskalen in jeweils zwei Gruppen zusammengefaßt wurden. So finden sich in Tabelle 6 für die EQ-Stufen unter 90, EQ unter 80 und unter 70 jeweils die erhaltenen Prozentwerte, die sich ergaben, wenn nur die Bewertung "abnorm" als testpositiv angesehen wird sowie wenn "abnorm" und "fraglich" als testpositiv gelten sollen.

Denver-Entwicklungs-skalen	Funktionelle Entwicklungsdiagnostik Entwicklungsquotient (EQ)				
	70	70 - 79	80 - 89	89	
abnorm	10	1	0	0	11
fraglich	0	0	2	1	3
normal	1	1	3	6	11
Summen	11	2	5	7	25

Tab. 5: Testergebnisse der Denver-Entwicklungsskalen und der funktionellen Entwicklungsdiagnostik bei Kindern unter drei Jahren (N = 25)

	Übereinstimmung		Korrektheit	
	positive	negative	positive	negative
1. EQ kleiner 70				
abnorm/fraglich + normal	91 %	93 %	91 %	93 %
abnorm + fraglich/ normal	91 %	71 %	71 %	91 %
2. EQ kleiner 80				
abnorm/fraglich + normal	84 %	100 %	100 %	86 %
abnorm + fraglich/ normal	84 %	75 %	79 %	82 %
3. EQ kleiner 90				
abnorm/fraglich + normal	61 %	100 %	100 %	50 %
abnorm + fraglich/ normal	72 %	86 %	93 %	55 %

Tab. 6: Übereinstimmung und Korrektheit der Ergebnisse der Denver-Entwicklungsskalen mit den Ergebnissen der funktionellen Entwicklungsdiagnostik bei Kindern unter drei Jahren (N = 25)

5.2. Gültigkeitsprüfung: Denver Entwicklungsskalen und Stanford-Binet Intelligenztest

In Tabelle 7 sind die Ergebnisse der 148 Kinder zwischen drei und sechs Jahren, die mit den Denver Entwicklungsskalen und dem Stanford-Binet Intelligenztest untersucht wurden, zusammengestellt. Der Anteil der Kinder mit einem IQ unter 70 liegt bei 15,5%, derer zwischen 70 und 80 bei 5%, derer zwischen 80 und 90 bei 10% und derjenigen mit einem IQ höher als 90 bei 69.5 %. Der Durchschnitts-IQ beträgt 104, die Standardabweichung 21,23. Die Prozentwerte der positiven und negativen Übereinstimmung bzw. Korrektheit finden sich in Tabelle 8 für jede der angegebenen IQ-Gruppen, wobei auch hier zunächst nur die Bewer-

tung "abnorm" als testpositiv betrachtet wurde und dann auch die Bewertung "fraglich" dazugenommen wurde.

Denver-Entwicklungs-skalen	Stanford-Binet Intelligenztest Intelligenzquotient (IQ)				
	70	70 - 79	80 - 89	89	
abnorm	21	6	9	11	47
fraglich	2	1	5	17	25
normal	0	0	1	75	76
Summen	23	7	15	103	148

Tab. 7: Testergebnisse der Denver-Entwicklungsskalen und des Stanford-Binet Intelligenztests von Kindern zwischen drei und sechs Jahren (N = 148)

	Übereinstimmung		Korrektheit	
	positive	negative	positive	negative
1. IQ kleiner 70				
abnorm/fraglich + normal	92 %	80 %	45 %	97 %
abnorm + fraglich/ normal	100 %	61 %	32 %	100 %
2. IQ kleiner 80				
abnorm/fraglich + normal	93 %	84 %	57 %	96 %
abnorm + fraglich/ normal	100 %	65 %	42 %	100 %
3. IQ kleiner 90				
abnorm/fraglich + normal	80 %	87 %	76 %	90 %
abnorm + fraglich/ normal	100 %	72 %	63 %	100 %

Tab. 8: Übereinstimmung und Korrektheit der Ergebnisse der Denver-Entwicklungsskalen mit den Ergebnissen des Stanford-Binet Intelligenztests bei Kindern zwischen drei und sechs Jahren (N = 148)

In Tabelle 9 finden sich die Ergebnisse der Untergruppe von 111 Kindern zwischen drei und sechs Jahren, die sowohl mit den Denver-Entwicklungsskalen und dem Stanford-Binet Test, als auch neurologisch untersucht wurden. Die Häufigkeit der IQ-Werte in dieser Gruppe ist mit der in der Hauptgruppe nahezu gleich, nämlich 16% der Kinder mit einem IQ unter 70, 5% zwischen 70 und 80, 10% zwischen 80 und 90 und 69% mit einem IQ gleich 90 bzw. höher.

47 Kinder, das entspricht 42%, waren neurologisch auffällig, 27 davon zeigten deutliche neurologische Ausfälle. Die häufigste Diagnose war cerebrale Bewegungsstörung unterschiedlichen Schweregrades.

In Tabelle 10 sind zunächst die Prozentwerte der positiven und negativen Übereinstimmung bzw. Korrektheit aufgeführt, wenn nur die Denver-Entwicklungsskalen und der Stanford-Binet Test bei dieser Untergruppe verglichen wurde.

Bei den Angaben in Tabelle 11 wurde zusätzlich die neurologische Beurteilung berücksichtigt, wobei ein Ergebnis der Denver-Entwicklungsskalen auch dann als richtig positiv gewertet wurde, wenn es zwar nicht mit dem Ergebnis des Entwicklungstests übereinstimmte, aber die neurologische Untersuchung das jeweilige Kind auffällig erscheinen ließ. Hier ist zu beachten, daß die Prozentwerte der negativen Übereinstimmung und der positiven Korrektheit im Vergleich zu Tabelle 10 zunehmen und zwar in jeder IQ-Gruppe und unabhängig davon, ob die"fraglichen" Ergebnisse der Denver-Entwicklungsskalen zu den "abnormen" oder zu den "normalen" Ergebnissen gezählt wurden.

Die Prozentwerte der positiven Übereinstimmung und der negativen Korrektheit nehmen im Vergleich zu Tabelle 10 ab, was darauf zurückzuführen ist, daß eine Anzahl von Kindern mit neurologischen Auffälligkeiten von den Denver-Entwicklungsskalen nicht als "abnorm" oder zumindest "fraglich" beurteilt wurden.

Denver-Entwicklungs-skalen	Stanford-Binet Intelligenztest und neurologische Untersuchung (+/-) Intelligenzquotient (IQ)								
	70		70-79		80-89		89		
	+	-	+	-	+	-	+	-	
abnorm	15	1	3	1	4	2	7	0	33
fraglich	2	0	0	1	2	3	8	0	16
normal	0	0	0	0	0	0	6	56	62
Summen	17	1	3	2	6	5	21	56	
	18		5		11		77		111

Tab. 9: Testergebnisse der Denver-Entwicklungsskalen und des Stanford-Binet Intelligenztests sowie die Ergebnisse der neurologischen Untersuchung (auffällig = +, unauffällig = -) von Kindern zwischen drei und sechs Jahren (N = 111)

	Übereinstimmung		Korrektheit	
	positive	negative	positive	negative
1. IQ kleiner 70				
abnorm/fraglich + normal	89 %	82 %	48 %	98 %
abnorm + fraglich/ normal	100 %	67 %	37 %	100 %
2. IQ kleiner 80				
abnorm/fraglich + normal	87 %	85 %	60 %	96 %
abnorm + fraglich/ normal	100 %	71 %	47 %	100 %
3. IQ kleiner 90				
abnorm/fraglich + normal	76 %	91 %	73 %	90 %
abnorm + fraglich/ normal	100 %	81 %	70 %	100 %

Tab. 10: Übereinstimmung und Korrektheit der Ergebnisse der Denver-Entwicklungsskalen mit den Ergebnissen des Stanford-Binet Intelligenztests bei Kindern zwischen drei und sechs Jahren (N = 111)

	Übereinstimmung		Korrektheit	
	positive	negative	positive	negative
1. IQ kleiner 70				
abnorm/fraglich + normal	62 %	95 %	91 %	76 %
abnorm + fraglich/ normal	87 %	89 %	86 %	90 %
2. IQ kleiner 80				
abnorm/fraglich + normal	62 %	97 %	94 %	76 %
abnorm + fraglich/ normal	88 %	92 %	90 %	90 %
3. IQ unter 90				
abnorm/fraglich + normal	60 %	100 %	100 %	72 %
abnorm + fraglich/ normal	89 %	100 %	100 %	90 %

Tab. 11: Übereinstimmung und Korrektheit der Denver-Entwicklungsskalen mit den Ergebnissen des Stanford-Binet Intelligenztests bei Kindern zwischen drei und sechs Jahren, die zusätzlich auch neurologisch untersucht wurden, unter Berücksichtigung der neurologischen Untersuchungsergebnisse (N = 111)

5.3 Stabilitätsprüfung

In Tabelle 12 sind die Ergebnisse von 38 Kindern zwischen drei und sechs Jahren zusammengefaßt, die im Abstand von höchstens zwei Wochen zweimal mit den Denver-Entwicklungsskalen getestet wurden. 18.5% der Kinder dieser Gruppe hatten einen IQ unter 70, 13% zwischen 70 und 80, 13% zwischen 80 und 90, bei 55% lag der IQ bei 90 oder darüber. Wurden die fraglichen Ergebnisse mit dem Ergebnis "abnorm" zusammengenommen, berechnete sich aus der Vierfeldertafel eine Stabilität von 95% und eine Wiederholbarkeit von 89%. Wurden die Kinder mit der Bewertung

"fraglich" zu denen mit der Bewertung "normal" gezählt, ergibt sich eine Übereinstimmung oder Stabilität von 95%, die Wiederholbarkeit liegt bei 86%.

		2. Test			
		abnorm	fraglich	normal	
1. Test	abnorm	12	1	0	13
	fraglich	1	2	2	4
	normal	0	0	20	20
		13	3	22	38

Tab. 12: Ergebnisse bei zweimaliger Testung mit den Denver-Entwicklungsskalen im Abstand von zwei Wochen bei Kindern zwischen drei und sechs Jahren (N = 38)

6. DISKUSSION

6.1 Die Validität der Denver-Entwicklungsskalen

Da die funktionelle Entwicklungsdiagnostik sich zum Zeitpunkt der Untersuchung noch in einem Stadium der Modifizierung befand, wurde nur eine relativ kleine Anzahl von Kindern mit den Denver-Entwicklungsskalen und der funktionellen Entwicklungsdiagnostik getestet, um erste Hinweise für einen umfangreicheren Vergleich zu erhalten.

Die positive bzw. negative Übereinstimmung von Denver-Entwicklungsskalen ist in der Erfassung von Kindern mit einem EQ von unter 70 mit 91% bzw. 93% sehr hoch. Nur 9% der gesuchten Kinder wurden mit den Denver-Entwicklungsskalen nicht gefunden und nur 9% der testpositiven Kinder hatten nach der funktionellen Entwicklungsdiagnostik einen EQ, der höher als 70 lag. Eine Verbesserung der Übereinstimmung wird nicht erreicht, wenn die Ergebnisse "abnorm" und "fraglich" als positiv zusammengefaßt werden, die positive Korrektheit verringert sich deutlich so daß bei der Beurteilung von Kindern mit einem EQ unter 70 die Denver-Beurteilung "abnorm" das beste Kriterium darstellen dürfte. Beim Screening von Kindern mit einem EQ unter 80 und unter 90 ergab sich eine geringere positive bzw. negative Übereinstimmung als bei Kindern mit einem EQ unter 70. Die positive Korrektheit bleibt hoch, wenn man für die Gruppe mit EQ unter 80 das Kriterium "abnorm" verwendet und bei der mit einem EQ unter 90 die Bewertungen "abnorm" und "fraglich" als testpositiv betrachtet.

Wie bei dem Vergleich der Ergebnisse der Denver-Entwicklungsskalen und der funktionellen Entwicklungsdiagnostik, so fällt auch beim Vergleich mit den Ergebnissen des Stanford-Binet Intelligenztests auf, daß bei sämtlichen drei IQ-Gruppen die positive Übereinstimmung mit 80 - 100% deutlich höher liegt als die negative mit Prozentwerten zwischen 61% und 84%, wie

aus Tabelle 8 zu ersehen ist. Ein ähnliches Verhältnis ergab sich auch bei der Gültigkeitsprüfung des Denver Developmental Screening Tests FRANKENBURG (1971). Anders ausgedrückt: Wird ein hoher bis sehr hoher Prozentsatz der Kinder mit einem IQ unter 90, unter 80 oder unter 70 von den Denver-Entwicklungsskalen richtig eingestuft und nicht als "normal" bewertet, womit eine wichtige Voraussetzung für einen brauchbaren Filtertest erfüllt ist. Es besteht jedoch ein nicht unerheblicher Anteil an falsch positiven Befunden. Das Verhältnis von richtig positiven zu falsch positiven Befunden beträgt zwischen 1 : 2 bzw. 1 : 0,3 entsprechend der positiven Korrektheit, die bei dieser Untersuchung erzielt wurde, mit Werten zwischen 32 % und 76 %.

Dieses Verhältnis wird durch die Prävalenz einer Krankheit in der untersuchten Gruppe beeinflußt (EGGSTEIN 1975), wie aus Abbildung 2 deutlich wird. Schon die um 1 % geringere Krankheitshäufigkeit der Darstellung zeigt ein deutliches Absinken des diagnostischen Aussagewertes oder was dasselbe ist, der positiven Korrektheit, bei gleichbleibender Spezifität und Empfindlichkeit des Tests. Wie leicht zu ersehen ist, hat der diagnostische Aussagewert eines Filtertests eine wesentliche Bedeutung für die Kosten-Nutzen-Relation eines Screeningprogrammes.

Da der Prozentsatz der Kinder mit einem unterdurchschnittlichen IQ bzw. EQ in dieser Untersuchungsgruppe (N = 148) über der Häufigkeit in der Gesamtbevölkerung liegt, muß dies bei der positiven bzw. negativen Übereinstimmung, die sich aus dieser Untersuchung berechnen ließ, berücksichtigt werden.

Würde man z.B. in einem Jahrgang eine Prävalenz von 2 % Kindern mit einem IQ unter 80 annehmen, so läge die positive Korrektheit der Denver-Entwicklungsskalen nur noch bei etwa 15 % unter Zufallsbedingungen, wie aus Abbildung 2 zu ersehen ist, wenn die positive und negative Übereinstimmung in diesem IQ-Bereich mit 93 % und 84 % entsprechend Tabelle 8 angesetzt wird. Das Verhältnis von richtig positiven zu falsch positiven Befunden wäre 1 : 5.

Eine wesentliche Verbesserung der positiven Korrektheit ist festzustellen, wenn man die in Tabelle 10 erstellten Werte betrachtet, bei der zusätzlich das Ergebnis der neurologischen Untersuchung berücksichtigt wurde. Die Untergruppe der 111 Kinder weicht in der Verteilung auf die vier IQ-Stufen nur geringfügig von der Verteilung in der Hauptgruppe von 148 Kindern ab, so daß die Ergebnisse beider Gruppen durchaus vergleichbar sind, wie sich auch aus der nahen identischen positiven bzw. negativen Übereinstimmung ergibt. Die Verbesserung der negativen Übereinstimmung auf Werte zwischen 89% und 100% bei Berücksichtigung der neurologischen Ergebnisse zeigt, daß mit den Merkmalen der Denver-Entwicklungsskalen auch Kinder mit Entwicklungsstörungen entdeckt werden, die mit einem Intelligenztest wie dem Stanford-Binet Intelligenztest nicht gemessen werden.

Da eine Untersuchungsmethode nach PRECHTL (1972) nur dann ergiebig sein kann, wenn ihre Komplexität in einem gewissen Verhältnis zur Komplexität des untersuchten Objekts steht, kann von einem Filtertest für Entwicklungs-screening keine differenzierte Diagnose erwartet werden. Unter Berücksichtigung dieser Problematik läßt sich sagen, daß mit den Denver-Entwicklungsskalen in einem hohen Prozentsatz Kinder mit auffälliger Entwicklung im Intelligenzbereich entdeckt werden, seine Gültigkeit für derartige Störungen somit deutlich wird.

Aus Tabelle 9 wird deutlich, daß die positive Übereinstimmung bei Berücksichtigung der neurologischen Untersuchungsergebnisse absinkt, also nicht alle neurologisch auffälligen Kinder von den Denver-Entwicklungsskalen erfaßt werden. Gründe für diese Veränderung sollen hier nicht weiter untersucht werden, weitere Arbeiten mit Analysen der Ursachen und einer möglichen Verbesserung der Denver-Entwicklungsskalen in dieser Hinsicht sind notwendig.

Ein weiterer Grund neben der Messung neurologischer Merkmale für die niedrige positive Korrektheit der Denver-Entwick-

lungsskalen bei der Suche nach Entwicklungsverzögerungen im Intelligenzbereich könnte darin liegen, daß die deutsche Standardisierung des Stanford-Binet Intelligenztests vor fast 20 Jahren durchgeführt wurde und die damals errechneten Altersnormen heute nicht mehr zutreffen. So ergeben sich bei den 1972 von THORNDIKE neu berechneten Stanford-Binet Skalen erhebliche Unterschiede zu den vorhergehenden anglo-amerikanischen Skalen, da Aufgaben bereits in einem jüngeren Alter gelöst werden. Bei Verwendung von Standardisierungsdaten für die deutsche Bearbeitung, die den heutigen Verhältnissen in Deutschland entsprechen, würden die erzielten IQ-Werte dieser Untersuchung niedriger ausfallen, und sich die Anzahl der von den Denver-Entwicklungsskalen als richtig "abnorm" oder "fraglich" Eingestuften erhöhen.

Stellt man sich die Frage, welche Beurteilung eines Kindes nach Testung mit den Denver-Entwicklungsskalen zur weiteren Diagnose Anlaß geben soll, ob lediglich Kinder mit der Bewertung "abnorm" oder auch diejenigen mit dem Ergebnis "fraglich" als auffällig anzusehen sind, ergeben sich aus Tabelle 8 einige Hinweise.

Betrachtet man zunächst die positive und negative Übereinstimmung bei der Erfassung von Kindern mit einem IQ unter 70, unter 80 und unter 90, so zeigt sich für jede IQ-Stufe bei Zusammenfassung der Beurteilungen "abnorm" und "fraglich" als testpositive Befunde eine höhere positive Übereinstimmung als bei der alleinigen Verwendung der Beurteilung "abnorm" als testpositiv.

Die negative Übereinstimmung und damit die positive Korrektheit verhält sich umgekehrt. Zur Entscheidung, welche der beiden Gültigkeitskriterien die positive Übereinstimmung oder die negative Übereinstimmung, für die Brauchbarkeit eines Filtertests von größerer Bedeutung ist, erweist sich Abb. 2 als hilfreich. Hieraus wird ersichtlich, daß bei einer Krankheitshäufigkeit, wie sie auch für die besprochenen Entwicklungsstörun-

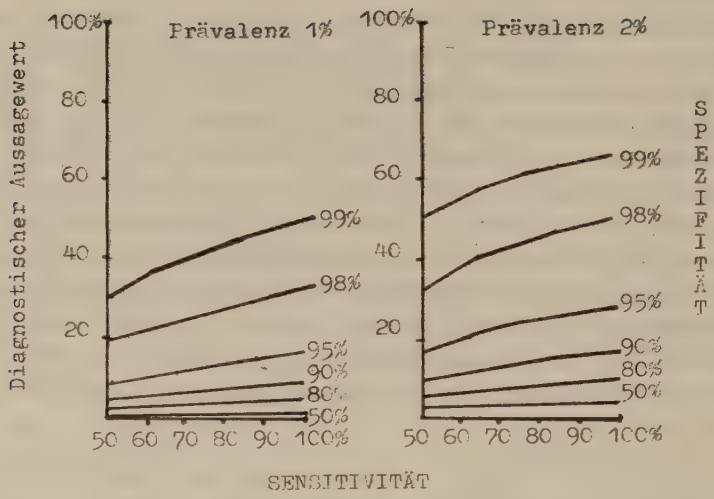

Abb. 1: Diagnostischer Aussagewert eines positiven Tests (nach EGGSTEIN 1975)

gen in etwa zutrifft, eine Minderung der negativen Übereinstimmung oder Spezifität eines Filtertests im Bereich der aus Tabelle 8 ersichtlichen Änderungen des Prozentwertes nach Verschiebung des als testpositiv geltenden Bewertungsrahmens, vor allem die positive Korrektheit oder den diagnostischen Wert eines Tests beeinflußt. Eine Minderung der positiven Übereinstimmung oder Empfindlichkeit im gleichen Umfange beeinträchtigt den diagnostischen Wert eines Tests nicht in diesem Ausmaße.

Bei der Suche nach Kindern mit IQ unter 70 und unter 80 erweist sich demnach die alleinige Verwendung der Beurteilung "abnorm" als testpositives Kriterium als sinnvoll. Zur Erfassung von Kindern mit einem IQ unter 90 erscheint auch die Einbeziehung der Bewertung "fraglich" als günstig, da hier sonst die positive Korrektheit zu sehr vermindert ist durch die zu niedrige positive Übereinstimmung. Der Wert der negativen Übereinstim-

mung bleibt hier unverändert bei 100%, wenn man berücksichtigt, daß auch neurologische Auffälligkeiten erfaßt wurden.

Eine diagnostische Zuordnung im Sinne eines IQ-Wertes durch die Beurteilungen "abnorm", "fraglich" und "normal" ist demnach nicht möglich. Es kann festgestellt werden, daß mit der Verwendung des Testurteils "abnorm" als testpositiv in hohem Maße Kinder mit einem IQ unter 70 und unter 80 mit den Denver-Entwicklungsskalen gefunden werden können, ohne eine unökonomisch hohe Zahl von falsch positiven Befunden. In noch höherem Maße gilt dies für Kinder mit einem IQ unter 90, wenn hierbei auch die Bewertung "fraglich" als testpositiv verwendet wird.

6.2 Die Reliabilität der Denver-Entwicklungsskalen

Die Prozentwerte der Wiederholbarkeit liegen mit 89% bzw. mit 86% deutlich über der untersten Grenze von 65%, die allgemein akzeptiert wird (LAASER 1976). Dieser Wert beschreibt vor allem die Stabilität der testpositiven Beurteilungen und erhält dann besondere Bedeutung, wenn die Häufigkeit der betreffenden Krankheit niedrig ist.

Das Maß der Stabilität, wie es oben definiert wurde, erwies sich mit 95% als sehr hoch.

7. ZUSAMMENFASSUNG

Die Denver-Entwicklungsskalen wurden hinsichtlich ihrer Eignung als Screeningtest für kindliche Entwicklungsstörungen überprüft.

Zur Untersuchung des Gütekriteriums Validität und um erste Hinweise für umfangreichere Studien zu erhalten, wurden 25 Kinder unter drei Jahren mit den Denver-Entwicklungsskalen und der Münchener funktionellen Entwicklungsdiagnostik getestet und die Ergebnisse verglichen. Es ergab sich eine hohe positive und negative Übereinstimmung.

Im Altersbereich von drei bis sechs Jahren wurden 148 Kinder mit den Denver-Entwicklungsskalen und dem Stanford-Binet Intelligenztest getestet, 111 Kinder dieser Gruppe wurden zusätzlich neurologisch untersucht. Auch hier ergab sich eine hohe positive und negative Übereinstimmung der Denver-Entwicklungsskalen und des Stanford-Binet Intelligenztests.

Wie zu erwarten, ließ sich die negative Übereinstimmung noch erhöhen, wenn die neurologischen Untersuchungsergebnisse berücksichtigt wurden. So lag die negative Übereinstimmung als wesentliches Maß für die Spezifität und damit die Effektivität einer Screeninguntersuchung bei 95% für die Erkennung von Kindern mit einem IQ unter 70, bei 97% und 100% für Kinder mit IQ kleiner 80 bzw. kleiner 90. Die positive Übereinstimmung als Maß für die Sensivität eines Tests lag zwischen 92% und 100%.

Insgesamt können die Denver-Entwicklungsskalen hinsichtlich ihrer Validität als geeigneter Screeningtest für Entwicklungsstörungen im Intelligenzbereich der Altersstufe von drei bis sechs Jahren angesehen werden. Eine Aussage über die Eignung in anderen Entwicklungsbereichen läßt sich aus dieser Untersuchung nicht ableiten und bedarf weiterer Prüfung.

Zur Untersuchung der Reliabilität der Denver-Entwicklungsskalen wurden 38 Kinder im Alter von drei bis sechs Jahren im Abstand von zwei Wochen getestet. Die Wiederholbarkeit lag zwischen 86% und 89%, die Stabilität war 95%.

Auch hinsichtlich der Zuverlässigkeit über einen bestimmten Zeitraum, können die Denver-Entwicklungsskalen demnach als geeigneter Screeningtest für den genannten Altersbereich gelten, wenn man den üblicherweise geforderten Gütemaßstab anlegt.

8. LITERATURVERZEICHNIS

1) BUCK, A.A., GART, J.J.:

Comparison of a screening test and a reference test in epidemilogical studies.
Am.J.Epidemiol 83: 586 - 592, 1966

2) EGGSTEIN, M.:

Bewertung von Laborbefunden in der Praxis.
Med.Welt 26: 474 - 478, 1975

3) EPSTEIN, F.H.:

Epidemilogical research in total communities.
Method.Inform.Med. 3: 18 - 22, 1964

4) FLEHMIG, I. et al.:

Denver-Entwicklungsskalen Testanweisung.
Hamburg 1973

5) FLEHMIG, I.:

Die Denver-Entwicklungsskalen (DES).
In: Frühe Hilfen - Wirksame Hilfen.
Schriftenreihe Lebenshilfe, Bd. 1, Marburg 1975

6) FRANKENBURG, W.K., DODDS, J.B.:

The Denver Developmental Screening Test.
J.Pediatr. 71: 181 - 191, 1967

7) FRANKENBURG, W.K., CAMP, B.W., VAN NALTA, P.A., DEMERSSEMAN, J.A., VORHEES, S.F.:

Reliability and Stability of the Denver Developmental Screening Test.
Child Develop. 42: 1315 - 1325, 1971

8) FRANKENBURG, W.K. et al.:

Validity of the Denver Developmental Screening Test.
Child Develop. 42: 475 - 486, 1971

9) FRANKENBURG, W.K., GOLDSTEIN, A.D., CAMP, B.W.:

The revised Denver Developmental Screening Test: Its accuracy as a screening instrument.
I. Pediatr. 79: 988 - 995, 1971

10) HELLBRÜGGE, Th.:
Über die Notwendigkeit der Vorsorgeuntersuchungen im Kleinkindesalter.
Deutsches Ärzteblatt 62: 1539 - 1546, 1965

11) HELLBRÜGGE, Th.:
Prävention und Rehabilitation in der frühen Kindheit.
Medizinische Klinik 66: 981 - 986, 1971

12) HELLBRÜGGE, Th., MENARA, D., SCHAMBERGER, R., STÜNKEL, S.:
Funktionelle Entwicklungsdiagnostik im zweiten Lebensjahr. FdM - Tabellen für die Praxis.
Fortschritte der Medizin 89: 558 - 562, 1971

13) HELLBRÜGGE, Th. (Hrsg.):
Fortschritte der Sozialpädiatrie 2. Hierin HELLBRÜGGE et al.: Funktionelle Entwicklungsdiagnostik für das erste Lebensjahr.
Urban und Schwarzenberg Verlag, München 1974

14) JAHN, E.:
Früherkennung chronischer Krankheiten als sozialmedizinische Aufgabe.
Bundesgesundheitsblatt 9: 217 - 228, 1966

15) KOLLER, S.:
Früherkennung von Krankheiten - Grundsätzliche Fragen.
In: Schröder, I., Immich, H.: Früherkennung von Krankheiten als methodisches Problem.
Verlag F.K. Schattauer, Stuttgart 1967

16) LAASER, U. et al.:
Prüfung eines empirisch bewährten Anamnesebogens auf Zuverlässigkeit.
Med.Welt 27: 2128 - 2134, 1976

17) LESSLER, K.:
Screening programs and the pediatricin.
Pediatrics 54: 608, 1974

18) LÜCKERT, R.:
Stanford-Binet Intelligenztest.
München 1956

19) OELLERICH, M., HAECKEL, R.:
Diagnostischer Wert klinisch-chemischer Untersuchungen.
Med.Welt 29: 866 - 872, 1978

20) PFLANZ, M.:
Allgemeine Epidemiologie.
Thieme Verlag, Stuttgart 1973

21) PRECHTL, H.F.:
Entwicklungsscreening: Von einem Dilemma ins andere.
Pädiatr.Praxis 11: 392 - 393, 1972

22) SCHAMBERGER, R.:
Funktionelle Entwicklungsdiagnostik und -therapie bei geistig behinderten und mehrfach behinderten Kindern.
Schriftenreihe Lebenshilfe Bd. 1. Marburg/Lahn 1975

23) TERMANN, C.M., MERILL, M.A.:
Stanford-Binet Intelligence Scale.
Cambridge 1960

24) THORNDIKE, R.L.:
Stanford-Binet Intelligence Scale, Third Revision, Form L-M, Columbia University 1972

25) TOUWEN, C.L.B., PRECHTL, H.F.R.:
The neurological examination of the child with minor nervous dysfunction.
Spastics Internat.Medical Pub. London 1970

26) VECCHIO, Th.J.:
New Engl.I.Med. 274: 1171 - 1173, 1966

27) WECHSLER, D.:
Die Messung der Intelligenz Erwachsener.
H.Huber Verlag Bern/Stuttgart 1956

28) WERNER, M.:
Med.Welt 28: 1254 - 1257, 1977

29) WILSON, I.M.C., JUNGNER, G.:
Principles and Practice of Screening for Disease.
Public Health Papers 34, WHO, 1968

Dem Vorstand des Instituts für Soziale Pädiatrie und Jugendmedizin, Herrn Professor Hellbrügge, danke ich für die Überlassung des Themas.

Herrn Professor Frankenburg, University of Colorado, Denver, danke ich für die Einführung in die Problematik, die Unterstützung bei der Durchführung und für seinen Rat bei der Diskussion der Ergebnisse.

Frau Dr.Orth von der Universitätskinderklinik München danke ich für die Überlassung der neurologischen Untersuchungsergebnisse.

Frau Dr.Lamby und Herrn Dr.Schirm danke ich für die Korrektur der Arbeit und ihre wertvollen Ratschläge.

LEBENSLAUF

Am 20.Januar 1950 wurde ich als ältester Sohn des Bauingenieurs Otto Roßberger und seiner Ehefrau, der Juristin Hanna Roßberger, geb. Merkenschlager, in München geboren.

Nach Besuch der Volksschule und des Humanistischen Gymnasiums München-Pasing legte ich das Abitur im Juni 1969 ab.

Ab dem Sommersemester 1970 immatrikulierte ich mich an der Universität München für das Studium der Humanmedizin. Die ärztliche Vorprüfung legte ich im Wintersemester 1973, das Staatsexamen im Wintersemester 1976 ab.

Meine Medizinalassistentenzeit absolvierte ich an der Chirurg.Abteilung des Kreiskrankenhauses Bad Aibling, der III. Med.Klinik (Prof. Frosch) Augsburg und der Frauenklinik Dr.Geisenhofer in München. Im Oktober 1977 begann ich als Assistenzarzt an der Inneren Abteilung des Kreiskrankenhauses Altötting.

Seit dem 2.August 1976 bin ich mit der Heilpädagogin Anna Maria Roßberger, geb. Bräutigam, verheiratet.

D R U C K : Arbeitsgruppe Psychologie
Pläntschweg 22, 8000 München 60